

Mineria de dades

Grau en Estadística de la UAB
Curs 2010-2011
Professor: Joan Valls Marsal

Programa de l'assignatura

1. Introducció a R
 - Qüestions bàsiques. Escalars, vectors, factors i matrius
 - Lectura i escriptura de fitxers de dades
 - Programació de funcions i gràfics
2. Mètodes basats en arbres
 - Arbres de classificació
 - Arbres de regressió
 - Poda d'arbres (*tree pruning*)
3. Tècniques de *Support Vector Machine* i discriminació flexible
 - *Support vector classifier*
 - *Support Vector Machines* i *Kernels*
 - Mètodes de discriminació flexible
4. Xarxes Neuronals
 - Models estadístics i xarxes neuronals
 - Ajust de xarxes neuronals: valors inicials i entrenament
5. Mineria de textos
 - Taules lèxiques i construcció del diccionari
 - Anàlisi de correspondències múltiples aplicada a la mineria de textos
6. Problemes $p \gg n$
 - Quan p és molt més gran que n
 - *False Discovery Rate*

Bibliografia

- T Hastie, R Tibshirani, J Friedman. (2009) *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, Springer, New York.
- L Torgo. (2010) *Data Mining with R. Learning with Case Studies*, Chapman & Hall, Miami.
- J Faraway. (2006) *Extending de Linear Model with R*, Chapman & Hall, Miami.
- J Faraway. (2005) *Linear Models with R*, Chapman & Hall, Boca Raton.
- W Venables, B Ripley. (2002) *Modern Applied Statistics with S-PLUS*, Springer, New York.

Pràctiques

Les classes pràctiques es realitzaran mitjançant el programa R.

Projecte d'anàlisi de dades

Els alumnes hauran de cercar un conjunt de dades de temàtica lliure, sobre el que aplicar alguna/es de les eines de mineria de dades i presentar els resultats obtinguts en un informe. Per avaluar-lo es tindrà en compte: l'aplicació correcta dels mètodes estadístics, l'originalitat i interès del problema, la claredat i brevetat en la presentació dels resultats i la presentació oral de l'informe mitjançant unes transparències.

Avaluació

L'avaluació de l'assignatura serà continuada, mitjançant exàmens parcials (40%), entrega de treballs i exercicis (40%) i assistència a classe amb participació activa (20%). Els alumnes que cursin l'assignatura de forma virtual, ho hauran d'indicar al professor a l'inici de curs i, en aquest cas, no s'avaluarà la participació a classe, però en canvi es podran presentar als exàmens parcials (60%) i entregar els treballs i exercicis (40%). Aquells alumnes que no superin la fase d'avaluació continuada tindran l'oportunitat de presentar-se a un únic examen i serà obligatori la presentació dels treballs.

Informació addicional

Les classes tindran lloc els dilluns de 15h a 19h (de 15h a 17h la part teòrica i de 17h a 19h la pràctica). Una part de les classes teòriques es realitzarà a l'aula d'ordinadors. Les consultes de l'assignatura es realitzaran de manera concertada amb el professor (jvalls@mat.uab.cat).